PCT/EP2004/012459

Europäisches Patentamt | European Patent Office | Office européen des brevets

REC'D 25 NOV 2004
WIPO PCT

BEST AVAILABLE COPY

Bescheinigung | Certificate | Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. | Patent application No. | Demande de brevet n°

03025340.5

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

Anmeldung Nr:
Application no.: 03025340.5
Demande no:

Anmeldetag:
Date of filing: 04.11.03
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Roche Diagnostics GmbH
Sandhofer Strasse 116
68305 Mannheim
ALLEMAGNE
F. HOFFMANN-LA ROCHE AG
124 Grenzacherstrasse
4070 Basel
SUISSE

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Method for distinguishing immunologically defined ALL subtypes

In Anspruch genommene Priorät(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

G01N33/48

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LU MC NL
PT RO SE SI SK TR LI

F. Hoffmann-La Roche
Roche Diagnostics GmbH

November 04, 2003
R62510EP BÖ/AMS

Method for distinguishing immunologically defined ALL subtypes

The present invention is directed to a method for distinguishing immunologically defined ALL subtypes by determining the expression level of selected marker genes.

Leukemias are classified into four different groups or types: acute myeloid (AML), acute lymphatic (ALL), chronic myeloid (CML) and chronic lymphatic leukemia (CLL). Within these groups, several subcategories can be identified further using a panel of standard techniques as described below. These different subcatgories in leukemias are associated with varying clinical outcome and therefore are the basis for different treatment strategies. The importance of highly specific classification may be illustrated in detail further for the AML as a very heterogeneous group of diseases. Effort is aimed at identifying biological entities and to distinguish and classify subgroups of AML which are associated with a favorable, intermediate or unfavorable prognosis, respectively. In 1976, the FAB classification was proposed by the French-American-British co-operative group which was based on cytomorphology and cytochemistry in order to separate AML subgroups according to the morphological appearance of blasts in the blood and bone marrow. In addition, it was recognized that genetic abnormalities occurring in the leukemic blast had a major impact on the morphological picture and even more on the prognosis. So far, the karyotype of the leukemic blasts is the most important independent prognostic factor regarding response to therapy as well as survival.

Usually, a combination of methods is necessary to obtain the most important information in leukemia diagnostics: Analysis of the morphology and cytochemistry of bone marrow blasts and peripheral blood cells is necessary to establish the diagnosis. In some cases the addition of immunophenotyping is mandatory to separate very undifferentiated AML from acute lymphoblastic leukemia and CLL. Leukemia subtypes investigated can be diagnosed

by cytomorphology alone, only if an expert reviews the smears. However, a genetic analysis based on chromosome analysis, fluorescence in situ hybridization or RT-PCR and immunophenotyping is required in order to assign all cases in to the right category. The aim of these techniques besides diagnosis is mainly to determine the prognosis of the leukemia. A major disadvantage of these methods, however, is that viable cells are necessary as the cells for genetic analysis have to divide in vitro in order to obtain metaphases for the analysis. Another problem is the long time of 72 hours from receipt of the material in the laboratory to obtain the result. Furthermore, great experience in preparation of chromosomes and even more in analyzing the karyotypes is required to obtain the correct result in at least 90% of cases. Using these techniques in combination, hematological malignancies in a first approach are separated into chronic myeloid leukemia (CML), chronic lymphatic (CLL), acute lymphoblastic (ALL), and acute myeloid leukemia (AML). Within the latter three disease entities several prognostically relevant subtypes have been established. As a second approach this further sub-classification is based mainly on genetic abnormalities of the leukemic blasts and clearly is associated with different prognoses.

The sub-classification of leukemias becomes increasingly important to guide therapy. The development of new, specific drugs and treatment approaches requires the identification of specific subtypes that may benefit from a distinct therapeutic protocol and, thus, can improve outcome of distinct subsets of leukemia. For example, the new therapeutic drug (STI571, Imatinib) inhibits the CML specific chimeric tyrosine kinase BCR-ABL generated from the genetic defect observed in CML, the BCR-ABL-rearrangement due to the translocation between chromosomes 3 and 22 (t(9;22) (q34; q11)). In patients treated with this new drug, the therapy response is dramatically higher as compared to all other drugs that had been used so far. Another example is the subtype of acute myeloid leukemia AML M3 and its variant M3v both with karyotype t(15;17)(q22; q11-12). The introduction of a new drug (all-trans retinoic acid - ATRA) has improved the outcome in this subgroup of patient from about 50% to 85 % long-term survivors. As it is mandatory for these patients suffering from these specific leukemia subtypes to be identified as fast as possible [illegible]

Thus, the technical problem underlying the present invention was to provide means for leukemia diagnostics which overcome at least some of the disadvantages of the prior art diagnostic methods, in particular encompassing the time-consuming and unreliable combination of different methods and which provides a rapid assay to unambigously distinguish one subtype from another, e.g. by genetic analysis.

According to Golub et al. (Science, 1999, 286, 531-7), gene expression profiles can be used for class prediction and discriminating AML from ALL samples. However, for the analysis of acute leukemias the selection of the two different subgroups was performed using exclusively morphologic-phenotypical criteria. This was only descriptive and does not provide deeper insights into the pathogenesis or the underlying biology of the leukemia. The approach reproduces only very basic knowledge of cytomorphology and intends to differentiate classes. The data is not sufficient to predict prognostically relevant cytogenetic aberrations.

Furthermore, the international application WO-A 03/039443 discloses marker genes the expression levels of which are characteristic for certain leukemia, e.g. AML subtypes and additionally discloses methods for differentiating between the subtype of AML cells by determining the expression profile of the disclosed marker genes. However, WO-A 03/039443 does not provide guidance which set of distinct genes discriminate between two subtypes and, as such, can be routineously taken in order to distinguish one ALL subtype from another.

The problem is solved by the present invention, which provides a method for distinguishing immunologically defined ALL subtypes Pro-B-ALL, c-ALL, Pre-B-ALL, c-ALL/Pre-B-ALL, mature B-ALL, precursor B-ALL, Pro-T-ALL, Pre-T-ALL, cortical T-ALL, mature T-ALL, and/or T-ALL in a sample, the method comprising determining the expression level of markers selected from the markers identifiable by their Affymetrix Identification Numbers (affy id) as defined in Tables 1 and or 2,

wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table 1.1

is indicative for the presence of ball when ball is distinguished from all other subtypes,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table 1.2

is indicative for the presence of cpre when cpre is distinguished from all other subtypes,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 6, 8, 9, 10, 12, 13, 14, 16, 17, 18, 22, 23, 24, 25, 30, 31, 34, 38, 40, 42, 43, 44, 46, 48, and/or 49, of Table 1.3 and/or

a higher expression of at least one polynucleotide defined by any of the numbers 4, 5, 7, 11, 15, 19, 20, 21, 26, 27, 28, 29, 32, 33, 35, 36, 37, 39, 41, 45, 47, and/or 50 of Table 1.3

is indicative for the presence of cpreh when cpreh is distinguished from all other subtypes,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 43, 44, 45, 46, 47, and/or 48, of Table 1.4, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 16, 22, 39, 49, and/or 50 of Table 1.4

is indicative for the presence of kort when kort is distinguished from all other subtypes,

is indicative for the presence of pret when pret is distinguished from all other subtypes,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 6, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 25, 26, 27, 28, 29, 32, 35, 36, 37, 38, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table 1.6, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 5, 7, 10, 20, 22, 23, 24, 30, 31, 33, 34, and/or 39 of Table 1.6,

is indicative for the presence of prob when prob is distinguished from from all other subtypes,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 3, 4, 8, 10, 12, 15, 17, 20, 23, 24, 25, 27, 28, 29, 30, 31, 34, 36, 37, 40, 42, 44, 45, 46, 49, and/or 50 of Table 2.1, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 2, 5, 6, 7, 9, 11, 13, 14, 16, 18, 19, 21, 22, 26, 32, 33, 35, 38, 39, 41, 43, 47, 48,

is indicative for the presence of ball when ball is distinguished from cpre,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table of Table 2.2, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 26, and/or 37, of Table 2.2

is indicative for the presence of ball when ball is distinguished from cpreph,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 30, 31, 33, 34, 36, 37, 38, 39, 40, 41, 42, 43, 45, 46, 47, 48, and/or 49, of Table 2.3, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 6, 7, 27, 29, 32, 35, 44, and/or 50 of Table 2.3

is indicative for the presence of ball when ball is distinguished from kort,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 3, 5, 6, 7, 13, 17, 18, 19, 21, 22, 26, 27, 30, 32, 34, 36, 38, 40, 47, and/or 48, of Table 2.4, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 1, 2, 4, 8, 9, 10, 11, 12, 14,15, 16, 20, 23, 24, 25, 28, 29, 31,33, 35,37, 39, 41, 42, 43, 44, 45, 46, 49, and/or 50 of Table 2.4

is indicative for the presence of ball when ball is distinguished from pret,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table of Table 2.5, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 29, 30 and/or 39, of Table 2.5,

is indicative for the presence of ball when ball is distinguished from prob,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 7, 9, 10, 11, 13, 17, 18, 21, 24, 25, 27, 29, 30, 31, 36, 37, 38, 40, 42, 43, 45, 46, 49, and/or 50 of Table 2.6, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 6, 8, 12, 14, 15, 16, 19, 20, 22, 23, 26, 28, 32, 33, 34, 35, 39, 41, 44, 47, and/or 48 of Table 2.6,

is indicative for the presence of cpre when cpre is distinguished from cpreph.

27, 28, 29, 30, 31, 32, 35, 36, 38, 40, 41, 43, 44, 45, 46, 48, 49, and/or 50 of Table 2.7, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 3, 7, 9, 11, 22, 26, 33, 34, 37, 39, 42, 47, of Table 2.7,

is indicative for cpre when cpre is distinguished from kort,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 20, 28, 31, 37, 38, and/or 50 of Table 2.8, and/or

a higher expression of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 29, 30, 32, 33, 34, 35, 36, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, and/or 49 of Table 2.8

is indicative for cpre when cpre is distinguished from pret,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 34, 35, 36, 37, 38, 39, 40, 42, 43, 44, 45, 46, 47, 48, and/or 50 of Table 2.9,

a higher expression of at least one polynucleotide defined by any of the numbers 26, 33, 41, and/or 49 of Table 2.9

is indicative for cpre when cpre is distinguished from prob,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 3, 6, 12, 17, 23, 28, 34, 35, and/or 41, of Table 2.10, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 1, 2, 4, 5, 7, 8, 9, 10, 11, 13, 14, 15, 16, 18, 19, 20, 21, 22, 24, 25, 26, 27, 29, 30, 31, 32, 33, 36, 37, 38, 39, 40, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table 2.10

is indicative for cpreph when cpreph is distinguished from kort,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 42, and/or 43, of Table 2.11, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 44, 45, 46, 47, 48, 49, and/or 50 of Table 2.11,

is indicative for cpreph when cpreph is distinguished from pret,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 3, 5, 8, 9, 11, 12, 13, 15, 18, 21, 24, 27, 28, 29, 32, 34, 36, 38, 41, 42, 43, 46, 47, 48, of Table 2.12, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 2, 4, 6, 7, 10, 14, 16, 17, 19, 20, 22, 23, 25, 26, 30, 31, 33, 35, 37, 39, 40, 44, 45, 49, and/or 50 of Table 2.12

is indicative for cpreph when cpreph is distinguished from prob

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 19, and/or 40, of Table 2.13

a higher expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table 2.13,

is indicative for kort when kort is distinguished from pret,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 4, 7, 9, 10, 11, 13, 14, 15, 16, 17, 20, 21, 22, 28, 29, 31, 32, 33, 35, 36, 37, 40, 41, 42, 43, 45, 47, 48, and/or 50 of Table 2.14, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 2, 3, 5, 6, 8, 12, 18, 19, 23, 24, 25, 26, 27, 30, 34, 38, 39, 44, 46, and/or 49 of Table 2.14

22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table 2.15,

is indicative for pret when pret is distinguished from prob.

As used herein, the following abbreviations represent the classified immunologically defined ALL subtypes:

ball=Mature B-ALL

cpre=c-ALL/Pre-B-ALL without t(9;22)

cpreph= c-ALL/Pre-B-ALL with t(9;22)

kort=Cortical T-ALL

pret=Pre-T-ALL

prob=Pro-B-ALL

According to the present invention, a "sample" means any biological material containing genetic information in the form of nucleic acids or proteins obtainable or obtained from an individual. The sample includes e.g. tissue samples, cell samples, bone marrow and/or body fluids such as blood, saliva, semen. Preferably, the sample is blood or bone marrow, more preferably the sample is bone marrow. The person skilled in the art is aware of methods, how to isolate nucleic acids and proteins from a sample. A general method for isolating and preparing nucleic acids from a sample is outlined in Example 3.

According to the present invention, the term "lower expression" is generally assigned to all by numbers and Affymetrix Id. definable polynucleotides the t-values and fold change (fc) values of which are negative, as indicated in the Tables. Accordingly, the term "higher expression" is generally assigned to all by numbers and Affymetrix Id. definable polynucleotides the t-values and fold change (fc) values of which are positive.

According to the present invention, the term "expression" refers to the process by which mRNA or a polypeptide is produced based on the nucleic acid sequence of a gene, i.e. „expression“ also includes the formation of mRNA upon transcription. In accordance with

the present invention, the term „determining the expression level" preferably refers to the determination of the level of expression, namely of the markers.

Generally, "marker" refers to any genetically controlled difference which can be used in the genetic analysis of a test versus a control sample, for the purpose of assigning the sample to a defined genotype or phenotype. As used herein, "markers" refer to genes which are differentially expressed in, e.g., different AML subtypes. The markers can be defined by their gene symbol name, their encoded protein name, their transcript identification number (cluster identification number), the data base accession number, public accession number or GenBank identifier or, as done in the present invention, Affymetrix identification number, chromosomal location, UniGene accession number and cluster type, LocusLink accession number (see Examples and Tables).

The Affymetrix identification number (affy id) is accessible for anyone and the person skilled in the art by entering the "gene expression omnibus" internet page of the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/geo/). In particular, the affy id's of the polynucleotides used for the method of the present invention are derived from the so-called U133 chip. The sequence data of each identification number can be viewed at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL96

Generally, the expression level of a marker is determined by the determining the expression of its corresponding "polynucleotide" as described hereinafter.

According to the present invention, the term „polynucleotide" refers, generally, to a DNA, in particular cDNA, or RNA, in particular a cRNA, or a portion thereof or a polypeptide or a portion thereof. In the case of RNA (or cDNA), the polynucleotide is formed upon transcription of a nucleotide sequence which is capable of expression. The polynucleotide fragments refer to fragments preferably of between at least 8, such as 10, 12, 15 or 18 nucleotides and at least 50, such as 60, 80, 100, 200 or 300 nucleotides in length, or a [illegible]

The determination of the expression level may be effected at the transcriptional or translational level, i.e. at the level of mRNA or at the protein level. Protein fragments such as peptides or polypeptides advantageously comprise between at least 6 and at least 25, such as 30, 40, 80, 100 or 200 consecutive amino acids representative of the corresponding full length protein. Six amino acids are generally recognized as the lowest peptidic stretch giving rise to a linear epitope recognized by an antibody, fragment or derivative thereof. Alternatively, the proteins or fragments thereof may be analysed using nucleic acid molecules specifically binding to three-dimensional structures (aptamers).

Depending on the nature of the polynucleotide or polypeptide, the determination of the expression levels may be effected by a variety of methods. For determining and detecting the expression level, it is preferred in the present invention that the polynucleotide, in particular the cRNA, is labelled.

The labelling of the polynucleotide or a polypeptide can occur by a variety of methods known to the skilled artisan. The label can be fluorescent, chemiluminescent, bioluminescent, radioactive (such as ^{3}H or ^{32}P). The labelling compound can be any labelling compound being suitable for the labelling of polynucleotides and/or polypeptides. Examples include fluorescent dyes, such as fluorescein, dichlorofluorescein, hexachlorofluorescein, BODIPY variants, ROX, tetramethylrhodamin, rhodamin X, Cyanine-2, Cyanine-3, Cyanine-5, Cyanine-7, IRD40, FluorX, Oregon Green, Alexa variants (available e.g. from Molecular Probes or Amersham Biosciences) and the like, biotin or biotinylated nucleotides, digoxigenin, radioisotopes, antibodies, enzymes and receptors. Depending on the type of labelling, the detection is done via fluorescence measurements, conjugation to streptavidin and/or avidin, antigen-antibody- and/or antibody-antibody-interactions, radioactivity measurements, as well as catalytic and/or receptor/ligand interactions. Suitable methods include the direct labelling (incorporation) method, the amino-modified (amino-allyl) nucleotide method (available e.g. from Ambion), and the primer tagging method (DNA dendrimer labelling, as kit available e.g. from Genisphere). Particularly preferred for the present invention is the use of biotin or biotinylated nucleotides for labelling, with the latter being directly incorporated into, e.g. the cRNA polynucleotide by in vitro transcription.

If the polynucleotide is mRNA, cDNA may be prepared into which a detectable label, as exemplified above, is incorporated. Said detectably labelled cDNA, in single-stranded form, may then be hybridised, preferably under stringent or highly stringent conditions to a panel of single-stranded oligonucleotides representing different genes and affixed to a solid support such as a chip. Upon applying appropriate washing steps, those cDNAs will be detected or quantitatively detected that have a counterpart in the oligonucleotide panel. Various advantageous embodiments of this general method are feasible. For example, the mRNA or the cDNA may be amplified e.g. by polymerase chain reaction, wherein it is preferable, for quantitative assessments, that the number of amplified copies corresponds relative to further amplified mRNAs or cDNAs to the number of mRNAs originally present in the cell. In a preferred embodiment of the present invention, the cDNAs are transcribed into cRNAs prior to the hybridisation step wherein only in the transcription step a label is incorporated into the nucleic acid and wherein the cRNA is employed for hybridisation. Alternatively, the label may be attached subsequent to the transcription step.

Similarly, proteins from a cell or tissue under investigation may be contacted with a panel of aptamers or of antibodies or fragments or derivatives thereof. The antibodies etc. may be affixed to a solid support such as a chip. Binding of proteins indicative of an AML subtype may be verified by binding to a detectably labelled secondary antibody or aptamer. For the labelling of antibodies, it is referred to Harlow and Lane, "Antibodies, a laboratory manual", CSH Press, 1988, Cold Spring Harbor. Specifically, a minimum set of proteins necessary for diagnosis of all AML subtypes may be selected for creation of a protein array system to make diagnosis on a protein lysate of a diagnostic bone marrow sample directly. Protein Array Systems for the detection of specific protein expression profiles already are available (for example: Bio-Plex, BIORAD, München, Germany). For this application preferably antibodies against the proteins have to be produced and immobilized on a platform e.g. glasslides or microtiterplates. The immobilized antibodies can be labelled with a reactant specific for the certain target proteins as discussed above. The reactants can include enzyme substrates, DNA, receptors, antigens or antibodies to create for example a capture sandwich immunoassay.

the q value is associated with each tested feature. The q value is similar to the p value, except it is a measure of significance in terms of the false discovery rate rather than the false positive rate (Storey JD and Tibshirani R. Proc.Natl.Acad.Sci., 2003, Vol. 100:9440-5.

In a preferred embodiment of the present invention, markers as defined in Tables 1.1-2.15 having a q-value of less than 3E-06, more preferred less than 1.5E-09, most preferred less than 1.5E-11, are measured.

Of the above defined markers, the expression level of at least two, preferably of at least ten, more preferably of at least 25, most preferably of 50 of at least one of the Tables of the markers is determined.

In another preferred embodiment, the expression level of at least 2, of at least 5, of at least 10 out of the markers having the numbers 1 – 10, 1-20, 1-40, 1-50 of at least one of the Tables 1.1-2.15 are measured.

The level of the expression of the „marker“, i.e. the expression of the polynucleotide is indicative of the ALL subtype of a cell or an organism. The level of expression of a marker or group of markers is measured and is compared with the level of expression of the same marker or the same group of markers from other cells or samples. The comparison may be effected in an actual experiment or in silico. When the expression level also referred to as expression pattern or expression signature (expression profile) is measurably different, there is according to the invention a meaningful difference in the level of expression. Preferably the difference at least is 5 %, 10% or 20%, more preferred at least 50% or may even be as high as 75% or 100%. More preferred the difference in the level of expression is at least 200%, i.e. two fold, at least 500%, i.e. five fold, or at least 1000%, i.e. 10 fold.

Accordingly, the expression level of markers expressed lower in a first subtype than in at least one second subtype, which differs from the first subtype, is at least 5 %, 10% or 20%, more preferred at least 50% or may even be 75% or 100%, i.e. 2-fold lower, preferably at least 10-fold, more preferably at least 50-fold, and most preferably at least 100-fold lower in the first subtype. On the other hand, the expression level of markers expressed higher in a first subtype than in at least one second subtype, which differs from the first subtype, is at least 5 %, 10% or 20%, more preferred at least 50% or may even be 75% or 100%, i.e.

2-fold higher, preferably at least 10-fold, more preferably at least 50-fold, and most preferably at least 100-fold higher in the first subtype.

In another embodiment of the present invention, the sample is derived from an individual having leukaemia, preferably ALL.

For the method of the present invention it is preferred if the polynucleotide the expression level of which is determined is in form of a transcribed polynucleotide. A particularly preferred transcribed polynucleotide is an mRNA, a cDNA and/or a cRNA, with the latter being preferred. Transcribed polynucleotides are isolated from a sample, reverse transcribed and/or amplified, and labelled, by employing methods well-known the person skilled in the art (see Example 3). In a preferred embodiment of the methods according to the invention, the step of determining the expression profile further comprises amplifying the transcribed polynucleotide.

In order to determine the expression level of the transcribed polynucleotide by the method of the present invention, it is preferred that the method comprises hybridizing the transcribed polynucleotide to a complementary polynucleotide, or a portion thereof, under stringent hybridization conditions, as described hereinafter.

The term "hybridizing" means hybridization under conventional hybridization conditions, preferably under stringent conditions as described, for example, in Sambrook, J., et al., in "Molecular Cloning: A Laboratory Manual" (1989), Eds. J. Sambrook, E. F. Fritsch and T. Maniatis, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY and the further definitions provided above. Such conditions are, for example, hybridization in 6x SSC, pH 7.0 / 0.1% SDS at about 45°C for 18-23 hours, followed by a washing step with 2x SSC/0.1% SDS at 50°C. In order to select the stringency, the salt concentration in the washing step can for example be chosen between 2x SSC/0.1% SDS at room temperature for low stringency and 0.2x SSC/0.1% SDS at 50°C for high stringency. In addition, the temperature of the washing step can be varied between room temperature, ca. 22°C, for low stringency and 65°C to 70° C for high stringency. Also contemplated are [illegible]

SSPE (20X SSPE = 3M NaCl; 0.2M NaH2PO4; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 mg/ml salmon sperm blocking DNA, followed by washes at 50°C with 1 X SSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5x SSC). Variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

"Complementary" and "complementarity", respectively, can be described by the percentage, i.e. proportion, of nucleotides which can form base pairs between two polynucleotide strands or within a specific region or domain of the two strands. Generally, complementary nucleotides are, according to the base pairing rules, adenine and thymine (or adenine and uracil), and cytosine and guanine. Complementarity may be partial, in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be a complete or total complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has effects on the efficiency and strength of hybridization between nucleic acid strands.

Two nucleic acid strands are considered to be 100% complementary to each other over a defined length if in a defined region all adenines of a first strand can pair with a thymine (or an uracil) of a second strand, all guanines of a first strand can pair with a cytosine of a second strand, all thymine (or uracils) of a first strand can pair with an adenine of a second strand, and all cytosines of a first strand can pair with a guanine of a second strand, and vice versa. According to the present invention, the degree of complementarity is determined over a stretch of 20, preferably 25, nucleotides, i.e. a 60% complementarity means that within a region of 20 nucleotides of two nucleic acid strands 12 nucleotides of the first strand can base pair with 12 nucleotides of the second strand according to the above ruling, either as a stretch of 12 contiguous nucleotides or interspersed by non-pairing nucleotides, when the two strands are attached to each other over said region of 20 nucleotides. The degree of complementarity can range from at least about 50% to full, i.e. 100% complementarity. Two single nucleic acid strands are said to be "substantially complementary" when they are at least about 80% complementary, preferably about 90% or higher. For carrying out the method of the present invention substantial complementarity is preferred.

Preferred methods for detection and quantification of the amount of polynucleotides, i.e. for the methods according to the invention allowing the determination of the level of expression of a marker, are those described by Sambrook et al. (1989) or real time methods known in the art as the TaqMan® method disclosed in WO92/02638 and the corresponding U.S. 5,210,015, U.S. 5,804,375, U.S. 5,487,972. This method exploits the exonuclease activity of a polymerase to generate a signal. In detail, the (at least one) target nucleic acid component is detected by a process comprising contacting the sample with an oligonucleotide containing a sequence complementary to a region of the target nucleic acid component and a labeled oligonucleotide containing a sequence complementary to a second region of the same target nucleic acid component sequence strand, but not including the nucleic acid sequence defined by the first oligonucleotide, to create a mixture of duplexes during hybridization conditions, wherein the duplexes comprise the target nucleic acid annealed to the first oligonucleotide and to the labeled oligonucleotide such that the 3'-end of the first oligonucleotide is adjacent to the 5'-end of the labeled oligonucleotide. Then this mixture is treated with a template-dependent nucleic acid polymerase having a 5' to 3' nuclease activity under conditions sufficient to permit the 5' to 3' nuclease activity of the polymerase to cleave the annealed, labeled oligonucleotide and release labeled fragments. The signal generated by the hydrolysis of the labeled oligonucleotide is detected and/ or measured. TaqMan® technology eliminates the need for a solid phase bound reaction complex to be formed and made detectable. Other methods include e.g. fluorescence resonance energy transfer between two adjacently hybridized probes as used in the LightCycler® format described in U.S. 6,174,670.

A preferred protocol if the marker, i.e. the polynucleotide, is in form of a transcribed nucleotide, is described in Example 3, where total RNA is isolated, cDNA and, subsequently, cRNA is synthesized and biotin is incorporated during the transcription reaction. The purified cRNA is applied to commercially available arrays which can be obtained e.g. from Affymetrix. The hybridized cRNA is detected according to the methods described in Example 3. The arrays are produced by photolithography or other methods known to experts skilled in the art e.g. from U.S. 5,445,934, U.S. 5,744,305, U.S. [illegible], U.S. [illegible] and EP [illegible] or EP [illegible], or as described hereinafter in [illegible].

expression level of the polynucleotides or polypeptides is detected using a compound which specifically binds to the polynucleotide of the polypeptide of the present invention.

As used herein, "specifically binding" means that the compound is capable of discriminating between two or more polynucleotides or polypeptides, i.e. it binds to the desired polynucleotide or polypeptide, but essentially does not bind unspecifically to a different polynucleotide or polypeptide.

The compound can be an antibody, or a fragment thereof, an enzyme, a so-called small molecule compound, a protein-scaffold, preferably an anticalin. In a preferred embodiment, the compound specifically binding to the polynucleotide or polypeptide is an antibody, or a fragment thereof.

As used herein, an "antibody" comprises monoclonal antibodies as first described by Köhler and Milstein in Nature 278 (1975), 495-497 as well as polyclonal antibodies, i.e. entibodies contained in a polyclonal antiserum. Monoclonal antibodies include those produced by transgenic mice. Fragments of antibodies include $F(ab')_2$, Fab and Fv fragments. Derivatives of antibodies include scFvs, chimeric and humanized antibodies. See, for example Harlow and Lane, loc. cit. For the detection of polypeptides using antibodies or fragments thereof, the person skilled in the art is aware of a variety of methods, all of which are included in the present invention. Examples include immunoprecipitation, Western blotting, Enzyme-linked immuno sorbent assay (ELISA), Enzyme-linked immuno sorbent assay (RIA), dissociation-enhanced lanthanide fluoro immuno assay (DELFIA), scintillation proximity assay (SPA). For detection, it is desirable if the antibody is labelled by one of the labelling compounds and methods described supra.

In another preferred embodiment of the present invention, the method for distinguishing immunologically defined ALL subtypes is carried out on an array.

In general, an "array" or "microarray" refers to a linear or two- or three dimensional arrangement of preferably discrete nucleic acid or polypeptide probes which comprises an intentionally created collection of nucleic acid or polypeptide probes of any length spotted onto a substrate/solid support. The person skilled in the art knows a collection of nucleic acids or polypeptide spotted onto a substrate/solid support also under the term "array". As known to the person skilled in the art, a microarray usually refers to a miniaturised array

arrangement, with the probes being attached to a density of at least about 10, 20, 50, 100 nucleic acid molecules referring to different or the same genes per cm^2. Furthermore, where appropriate an array can be referred to as "gene chip". The array itself can have different formats, e.g. libraries of soluble probes or libraries of probes tethered to resin beads, silica chips, or other solid supports.

The process of array fabrication is well-known to the person skilled in the art. In the following, the process for preparing a nucleic acid array is described. Commonly, the process comprises preparing a glass (or other) slide (e.g. chemical treatment of the glass to enhance binding of the nucleic acid probes to the glass surface), obtaining DNA sequences representing genes of a genome of interest, and spotting sequences these sequences of interest onto glass slide. Sequences of interest can be obtained via creating a cDNA library from an mRNA source or by using publicly available databases, such as GeneBank, to annotate the sequence information of custom cDNA libraries or to identify cDNA clones from previously prepared libraries. Generally, it is recommendable to amplify obtained sequences by PCR in order to have sufficient amounts of DNA to print on the array. The liquid containing the amplified probes can be deposited on the array by using a set of microspotting pins. Ideally, the amount deposited should be uniform. The process can further include UV-crosslinking in order to enhance immobilization of the probes on the array.

In a preferred embodiment, the array is a high density oligonucleotide (oligo) array using a light-directed chemical synthesis process, employing the so-called photolithography technology. Unlike common cDNA arrays, oligo arrays (according to the Affymetrix technology) use a single-dye technology. Given the sequence information of the markers, the sequence can be synthesized directly onto the array, thus, bypassing the need for physical intermediates, such as PCR products, required for making cDNA arrays. For this purpose, the marker, or partial sequences thereof, can be represented by 14 to 20 features, preferably by less than 14 features, more preferably less than 10 features, even more preferably by 6 features or less, with each feature being a short sequence of nucleotides (oligonucleotide), which is a perfect match (PM) to a segment of the respective gene. The [illegible]

Advantageously, the method of the present invention is carried out in a robotics system including robotic plating and a robotic liquid transfer system, e.g. using microfluidics, i.e. channelled structured.

A particular preferred method according to the present invention is as follows:
1. Obtaining a sample, e.g. bone marrow or peripheral blood aliquots, from a patient having ALL
2. Extracting RNA, preferably mRNA, from the sample
3. Reverse transcribing the RNA into cDNA
4. In vitro transcribing the cDNA into cRNA
5. Fragmenting the cRNA
6. Hybridizing the fragmented cRNA on standard microarrays
7. Determining hybridization

In another embodiment, the present invention is directed to the use of at least one marker selected from the markers identifiable by their Affymetrix Identification Numbers (affy id) as defined in Tables 1, and/or 2 for the manufacturing of a diagnostic for distinguishing immunologically defined ALL subtypes. The use of the present invention is particularly advantageous for distinguishing immunologically defined ALL subtypes in an individual having ALL. The use of said markers for diagnosis of immunologically defined leukemia subtypes, preferably based on microarray technology, offers the following advantages: (1) more rapid and more precise diagnosis, (2) easy to use in laboratories without specialized experience, (3) abolishes the requirement for analyzing viable cells for chromosome analysis (transport problem), and (4) very experienced hematologists for cytomorphology and cytochemistry, immunophenotyping as well as cytogeneticists and molecularbiologists are no longer required.

Accordingly, the present invention refers to a diagnostic kit containing at least one marker selected from the markers identifiable by their Affymetrix Identification Numbers (affy id) as defined in Tables 1, and/or 2 for distinguishing immunologically defined ALL subtypes, in combination with suitable auxiliaries. Suitable auxiliaries, as used herein, include buffers, enzymes, labelling compounds, and the like. In a preferred embodiment, the marker contained in the kit is a nucleic acid molecule which is capable of hybridizing to the mRNA corresponding to at least one marker of the present invention. Preferably, the at least one nucleic acid molecule is attached to a solid support, e.g. a polystyrene microtiter dish, nitrocellulose membrane, glass surface or to non-immobilized particles in solution.

In another preferred embodiment, the diagnostic kit contains at least one reference for a Pro-B-ALL, c-ALL, Pre-B-ALL, c-ALL/Pre-B-ALL, mature B-ALL, precursor B-ALL, Pro-T-ALL, Pre-T-ALL, cortical T-ALL, mature T-ALL, and/or T-ALL subtype. As used herein, the reference can be a sample or a data bank.

In another embodiment, the present invention is directed to an apparatus for distinguishing immunologically defined AML subtypes subtypes Pro-B-ALL, c-ALL, Pre-B-ALL, c-ALL/Pre-B-ALL, mature B-ALL, precursor B-ALL, Pro-T-ALL, Pre-T-ALL, cortical T-ALL, mature T-ALL, and/or T-ALL in a sample, containing a reference data bank obtainable by comprising

(a) compiling a gene expression profile of a patient sample by determining the expression level at least one marker selected from the markers identifiable by their Affymetrix Identification Numbers (affy id) as defined in Tables 1, and/or 2, and

(b) classifying the gene expression profile by means of a machine learning algorithm.

According to the present invention, the "machine learning algorithm" is a computational-based prediction methodology, also known to the person skilled in the art as "classifier", employed for characterizing a gene expression profile. The signals corresponding to a certain expression level which are obtained by the microarray hybridization are subjected to the algorithm in order to classify the expression profile. Supervised learning involves "training" a classifier to recognize the distinctions among classes and then "testing" the accuracy of the classifier on an independent test set. For new, unknown sample the classifier shall predict into which class the sample belongs.

Preferably, the machine learning algorithm is selected from the group consisting of Weighted Voting, K-Nearest Neighbors, Decision Tree Induction, Support Vector Machines (SVM), and Feed-Forward Neural Networks. Most preferably, the machine learning algorithm is Support Vector Machine, such as polynomial kernel and Gaussian Radial Basis Function-kernel SVM models.

SVC, linear kernel (http://www.csie.ntu.edu.tw/~cjlin/libsvm/)). The skilled artisan is furthermore referred to Brown et al., Proc.Natl.Acad.Sci., 2000; 97: 262-267, Furey et al., Bioinformatics. 2000; 16: 906-914, and Vapnik V. Statistical Learning Theory. New York: Wiley, 1998.

In detail, the classification accuracy of a given gene list for a set of microarray experiments can be estimated using Support Vector Machines (SVM) as supervised learning technique. Generally, SVMs are trained using differentially expressed genes which were identified on a subset of the data and then this trained model is employed to assign new samples to those trained groups from a second and different data set. Differentially expressed genes were identified applying ANOVA and t-test-statistics (Welch t-test). Based on identified distinct gene expression signatures respective training sets consisting of 2/3 of cases and test sets with 1/3 of cases to assess classification accuracies are designated. Assignment of cases to training and test set is randomized and balanced by diagnosis. Based on the training set a Support Vector Machine (SVM) model is built.

According to the present invention, the apparent accuracy, i.e. the overall rate of correct predictions of the complete data set was estimated by 10fold cross validation. This means that the data set was divided into 10 approximately equally sized subsets, an SVM-model was trained for 9 subsets and predictions were generated for the remaining subset. This training and prediction process was repeated 10 times to include predictions for each subset. Subsequently the data set was split into a training set, consisting of two thirds of the samples, and a test set with the remaining one third. Apparent accuracy for the training set was estimated by 10fold cross validation (analogous to apparent accuracy for complete set). A SVM-model of the training set was built to predict diagnosis in the independent test set, thereby estimating true accuracy of the prediction model. This prediction approach was applied both for overall classification (multi-class) and binary classification (diagnosis X => yes or no). For the latter, sensitivity and specificity were calculated:

Sensitivity = (number of positive samples predicted)/(number of true positives)

Specificity = (number of negative samples predicted)/(number of true negatives)

In a preferred embodiment, the reference data bank is backed up on a computational data memory chip which can be inserted in as well as removed from the apparatus of the present

invention, e.g. like an interchangeable module, in order to use another data memory chip containing a different reference data bank.

The apparatus of the present invention containing a desired reference data bank can be used in a way such that an unknown sample is, first, subjected to gene expression profiling, e.g. by microarray analysis in a manner as described supra or in the art, and the expression level data obtained by the analysis are, second, fed into the apparatus and compared with the data of the reference data bank obtainable by the above method. For this purpose, the apparatus suitably contains a device for entering the expression level of the data, for example a control panel such as a keyboard. The results, whether and how the data of the unknown sample fit into the reference data bank can be made visible on a provided monitor or display screen and, if desired, printed out on an incorporated of connected printer.

Alternatively, the apparatus of the present invention is equipped with particular appliances suitable for detecting and measuring the expression profile data and, subsequently, proceeding with the comparison with the reference data bank. In this embodiment, the apparatus of the present invention can contain a gripper arm and/or a tray which takes up the microarray containing the hybridized nucleic acids.

In another embodiment, the present invention refers to a reference data bank for distinguishing immunologically defined ALL subtypes Pro-B-ALL, c-ALL, Pre-B-ALL, c-ALL/Pre-B-ALL, mature B-ALL, precursor B-ALL, Pro-T-ALL, Pre-T-ALL, cortical T-ALL, mature T-ALL, and/or T-ALL in a sample obtainable by comprising

(a) compiling a gene expression profile of a patient sample by determining the expression level of at least one marker selected from the markers identifiable by their Affymetrix Identification Numbers (affy id) as defined in Tables 1, and/or and

(b) classifying the gene expression profile by means of a machine learning algorithm.

Preferably, the [illegible] is [illegible] in a computational [illegible]

The invention is further illustrated in the following table and examples, without limiting the scope of the invention:

TABLES 1.1-2.15

Tables 1.1-2.15 show ALLL subtype analysis of subtypes Pro-B-ALL, c-ALL, Pre-B-ALL, c-ALL/Pre-B-ALL, mature B-ALL, precursor B-ALL, Pro-T-ALL, Pre-T-ALL, cortical T-ALL, mature T-ALL, and/or T-ALL. The analysed markers are ordered according to their q-values, beginning with the lowest q-values.
For convenience and a better understanding, Tables 1.1 to 2.15 are accompanied with explanatory tables (Table 1.1A to 2.15A) where the numbering and the Affymetrix Id are further defined by other parameters, e.g. gene bank accession number.

EXAMPLES

Example 1: General experimental design of the invention and results

Acute lymphoblastic leukemia (ALL) is a heterogeneous group of diseases which are classified immunologically. Most of the clinically relevant subgroups are characterized by specific genetic translocations, i.e. translocations involving MLL (tMLL) in Pro-B-ALL, t(9;22) in c-ALL and Pre-B-ALL, and t(8;14) in mature B-ALL. While in childhood ALL gene expression profiling revealed specific gene signatures in cytogenetically defined subgroup the respective data are scarce in adult ALL and, in particular, it is not known if the immunologically defined subtypes of ALL which lack specific cytogenetic aberrations display a characteristic gene expression profile. We analyzed global gene expression signatures in bone marrow samples from 95 patients with newly diagnosed ALL by use of microarray technology (Pro-B-ALL n=18, c-ALL n=18, Pre-B-ALL n=5, c-ALL/Pre-B-ALL n=12, mature B-ALL n=11, precursor B-ALL n=3, Pro-T-ALL n=2, Pre-T-ALL n=8, cortical T-ALL n=14, mature T-ALL n=2, T-ALL n=2). The diagnosis was based on cytomorphology, immunophenotyping, and cytogenetic and molecular genetic analyses. All samples were hybridized onto U133 set microarrays (Affymetrix) representing >30,000 human transcripts. Differentially expressed genes were identified applying ANOVA and t-test-statistics (Welch ttest). To assess the false discovery rate we calculated q-values according to Storey et al., PNAS 2003. Moreover, based on identified distinct gene expression signatures we designated respective training sets consisting of 2/3 of cases and test samples with 1/3 of cases to assess classification accuracies. Assignment of cases to

training and test set was randomized and balanced by diagnosis. Based on the training set we built a Support Vector Machine (SVM) model. Classification accuracy was assessed in the test set. In a first step, precursor B-ALL and precursor T-ALL were distinguished in 31 independent test samples with an accuracy of 100%. In a second step samples were separated according to the EGIL classification (Pro-B-ALL, c-ALL, Pre-B-ALL, mature B-ALL, Pre-T-ALL, cortical T-ALL). Out of the 25 test samples 20 were classified correctly (accuracy: 80%). Samples misclassified were: c-ALL as Pre-B-ALL (n=2), c-ALL as mature B-ALL, cortical T-ALL as Pre-T-ALL, and Pre-B-ALL as mature B-ALL (one each). Samples with c-ALL and Pre-B-ALL were then further subgrouped genetically according to positivity/negativity for t(9;22). Out of 29 test samples 24 were classified correctly (accuracy: 83%). Sample misclassified were: c-ALL/Pre-B-ALL without t(9;22) as Pro-B-ALL and mature B-ALL (one each), c-ALL/Pre-B-ALL with t(9;22) as c-ALL/Pre-B-ALL without t(9;22) and mature B-ALL (one each), Pre-T-ALL as cortical T-ALL. These data demonstrate that distinct immunologically defined subtypes of ALL are characterized by specific gene expression profiles. Distinction between Tlineage and B-lineage disease is accomplished with 100% accuracy while misclassification occurs in cases belonging to subtypes closely related to each other with regard to the maturation status. Gene expression profiling of ALL may help to optimize diagnostics of ALL and to allow further insights into the pathogenesis of the biologically defined subgroups.

Example 2: General materials, methods and definitions of functional annotations

The methods section contains both information on statistical analyses used for identification of differentially expressed genes and detailed annotation data of identified microarray probesets.

<u>Affymetrix Probeset Annotation</u>

All annotation data of GeneChip® arrays are extracted from the NetAffx™ Analysis Center (internet website: www.affymetrix.com). Files for U133 set arrays, including U133A and U133B microarrays are derived from the June 2003 release. The original publication refers to: Liu G, Loraine AE, Shigeta R, Cline M, Cheng J, Valmeekam V, Sun [illegible]

chromomal location and functional annotation of the respective gene products. Sequence data are available for download in the NetAffx Download Center (www.affymetrix.com)

Data fields:
In the following section, the content of each field of the data files are described. Microarray probesets, for example found to be differentially expressed between different types of leukemia samples are further described by additional information. The fields are of the following types:

1. GeneChip Array Information
2. Probe Design Information
3. Public Domain and Genomic References

1. GeneChip Array Information

HG-U133 ProbeSet_ID:
HG-U133 ProbeSet_ID describes the probe set identifier. Examples are: 200007_at, 200011_s_at, 200012_x_at.

GeneChip:
The description of the GeneChip probe array name where the respective probeset is represented. Examples are: Affymetrix Human Genome U133A Array or Affymetrix Human Genome U133B Array.

2. Probe Design Information

Sequence Type:
The Sequence Type indicates whether the sequence is an Exemplar, Consensus or Control sequence. An Exemplar is a single nucleotide sequence taken directly from a public database. This sequence could be an mRNA or EST. A Consensus sequence, is a nucleotide sequence assembled by Affymetrix, based on one or more sequence taken from a public database.

Transcript ID:
The cluster identification number with a sub-cluster identifier appended.

Sequence Derived From:

The accession number of the single sequence, or representative sequence on which the probe set is based. Refer to the "Sequence Source" field to determine the database used.

Sequence ID:
For Exemplar sequences: Public accession number or GenBank identifier. For Consensus sequences: Affymetrix identification number or public accession number.

Sequence Source:
The database from which the sequence used to design this probe set was taken. Examples are: GenBank®, RefSeq, UniGene, TIGR (annotations from The Institute for Genomic Research).

3. Public Domain and Genomic References

Most of the data in this section come from LocusLink and UniGene databases, and are annotations of the reference sequence on which the probe set is modeled.

Gene Symbol and Title:
A gene symbol and a short title, when one is available. Such symbols are assigned by different organizations for different species. Affymetrix annotational data come from the UniGene record. There is no indication which species-specific databank was used, but some of the possibilities include for example HUGO: The Human Genome Organization.

MapLocation:
The map location describes the chromosomal location when one is available.

Unigene_Accession:
UniGene accession number and cluster type. Cluster type can be "full length" or "est", or "---" if unknown.

Example 3: Sample preparation, processing and data analysis

Method 1:

Microarray analyses were performed utilizing the GeneChip® System (Affymetrix, Santa Clara, USA). Hybridization target preparations were performed according to recommended protocols (Affymetrix Technical Manual). In detail, at time of diagnosis, mononuclear cells were purified by Ficoll-Hypaque density centrifugation. They had been lysed immediately in RLT buffer (Qiagen, Hilden, Germany), frozen, and stored at -80°C from 1 week to 38 months. For gene expression profiling cell lysates of the leukemia samples were thawed, homogenized (QIAshredder, Qiagen), and total RNA was extracted (RNeasy Mini Kit, Qiagen). Subsequently, 5-10 µg total RNA isolated from 1 x 10^7 cells was used as starting material for cDNA synthesis with oligo[$(dT)_{24}$T7promotor$]_{65}$ primer (cDNA Synthesis System, Roche Applied Science, Mannheim, Germany). cDNA products were purified by phenol/chlorophorm/IAA extraction (Ambion, Austin, USA) and acetate/ethanol-precipitated overnight. For detection of the hybridized target nucleic acid biotin-labeled ribonucleotides were incorporated during the following *in vitro* transcription reaction (Enzo BioArray HighYield RNA Transcript Labeling Kit, Enzo Diagnostics). After quantification by spectrophotometric measurements and 260/280 absorbance values assessment for quality control of the purified cRNA (RNeasy Mini Kit, Qiagen), 15 µg cRNA was fragmented by alkaline treatment (200 mM Tris-acetate, pH 8.2/500 mM potassium acetate/150 mM magnesium acetate) and added to the hybridization cocktail sufficient for five hybridizations on standard GeneChip microarrays (300 µl final volume). Washing and staining of the probe arrays was performed according to the recommended Fluidics Station protocol (EukGE-WS2v4). Affymetrix Microarray Suite software (version 5.0.1) extracted fluorescence signal intensities from each feature on the microarrays as detected by confocal laser scanning according to the manufacturer's recommendations.

Expression analysis quality assessment parameters included visiual array inspection of the scanned image for the presence of image artifacts and correct grid alignment for the identification of distinct probe cells as well as both low 3'/5' ratio of housekeeping controls (mean: 1.90 for GAPDH) and high percentage of detection calls (mean: 46.3% present called genes). The 3' to 5' ratio of GAPDH probesets can be used to assess RNA sample and assay quality. Signal values of the 3' probe sets for GAPDH are compared to the Signal values of the corresponding 5' probe set. The ratio of the 3' probe set to the 5'

probe set is generally no more than 3.0. A high 3' to 5' ratio may indicate degraded RNA or inefficient synthesis of ds cDNA or biotinylated cRNA (GeneChip® Expression Analysis Technical Manual, www.affymetrix.com). Detection calls are used to determine whether the transcript of a gene is detected (present) or undetected (absent) and were calculated using default parameters of the Microarray Analysis Suite MAS 5.0 software package.

Method 2:

Bone marrow (BM) aspirates are taken at the time of the initial diagnostic biopsy and remaining material is immediately lysed in RLT buffer (Qiagen), frozen and stored at -80 C until preparation for gene expression analysis. For microarray analysis the GeneChip System (Affymetrix, Santa Clara, CA, USA) is used. The targets for GeneChip analysis are prepared according to the current Expression Analysis. Briefly, frozen lysates of the leukemia samples are thawed, homogenized (QIAshredder, Qiagen) and total RNA extracted (RNeasy Mini Kit, Qiagen).Normally 10 ug total RNA isolated from 1 x 107 cells is used as starting material in the subsequent cDNA-Synthesis using Oligo-dT-T7-Promotor Primer (cDNA synthesis Kit, Roche Molecular Biochemicals). The cDNA is purified by phenol-chlorophorm extraction and precipitated with 100% Ethanol over night. For detection of the hybridized target nucleic acid biotin-labeled ribonucleotides are incorporated during the in vitro transcription reaction (Enzo® BioArray™ HighYield™ RNA Transcript Labeling Kit, ENZO). After quantification of the purified cRNA (RNeasy Mini Kit, Qiagen), 15 ug are fragmented by alkaline treatment (200 mM Tris-acetate, pH 8.2, 500 mM potassium acetate, 150 mM magnesium acetate) and added to the hybridization cocktail sufficient for 5 hybridizations on standard GeneChip microarrays. Before expression profiling Test3 Probe Arrays (Affymetrix) are chosen for monitoring of the integrity of the cRNA. Only labeled cRNA-cocktails which showed a ratio of the messured intensity of the 3' to the 5' end of the GAPDH gene less than 3.0 are selected for subsequent hybridization on HG-U133 probe arrays (Affymetrix). Washing and staining the Probe arrays is performed as described (siehe Affymetrix-Original-Literatur (LOCKHART und LIPSHUTZ). The Affymetrix software (Microarray Suite, Version 4.0.1) extracted fluorescence intensities from each element on the arrays as detected by [illegible]

EPO - Munich
28
- 4. Nov. 2003

F. Hoffmann-La Roche AG
Roche Diagnostics GmbH

November 04, 2003
R62510EP BÖ/AMS

Claims

1. A method for distinguishing immunologically defined ALL subtypes Pro-B-ALL, c-ALL, Pre-B-ALL, c-ALL/Pre-B-ALL, mature B-ALL, precursor B-ALL, Pro-T-ALL, Pre-T-ALL, cortical T-ALL, mature T-ALL, and/or T-ALL in a sample, the method comprising determining the expression level of markers selected from the markers identifiable by their Affymetrix Identification Numbers (affy id) as defined in Tables 1 and/or 2,

wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table 1.1

is indicative for the presence of ball when ball is distinguished from all other subtypes,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table 1.2

is indicative for the presence of cpre when cpre is distinguished from all other subtypes,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 6, 8, 9, 10, 12, 13, 14, 16, 17, 18, 22, 23, 24, 25, 30, 31, 34, 38, 40, 42, 43, 44, 46, 48, and/or 49, of Table 1.3 and/or

a higher expression of at least one polynucleotide defined by any of the numbers 4, 5, 7, 11, 15, 19, 20, 21, 26, 27, 28, 29, 32, 33, 35, 36, 37, 39, 41, 45, 47, and/or 50 of Table 1.3

is indicative for the presence of cpreh when cpreh is distinguished from all other subtypes,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 43, 44, 45, 46, 47, and/or 48, of Table 1.4, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 16, 22, 39, 49, and/or 50 of Table 1.4

is indicative for the presence of kort when kort is distinguished from all other subtypes,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table 1.5

is indicative for the presence of pret when pret is distinguished from all other subtypes,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 6, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 25, 26, 27, 28, 29, 32, 35, 36, 37, 38, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table 1.6, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 5, 7, 10, 20, 22, 23, 24, 30, 31, 33, 34, and/or 39 of Table 1.6,

is indicative for the presence of prob when prob is distinguished from from all other subtypes,

a higher expression of at least one polynucleotide defined by any of the numbers 2, 5, 6, 7, 9, 11, 13, 14, 16, 18, 19, 21, 22, 26, 32, 33, 35, 38, 39, 41, 43, 47, 48,

is indicative for the presence of ball when ball is distinguished from cpre,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table of Table 2.2, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 26, and/or 37, of Table 2.2

is indicative for the presence of ball when ball is distinguished from cpreph,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28, 30, 31, 33, 34, 36, 37, 38, 39, 40, 41, 42, 43, 45, 46, 47, 48, and/or 49, of Table 2.3, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 6, 7, 27, 29, 32, 35, 44, and/or 50 of Table 2.3

is indicative for the presence of ball when ball is distinguished from kort,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 3, 5, 6, 7, 13, 17, 18, 19, 21, 22, 26, 27, 30, 32, 34, 36, 38, 40, 47, and/or 48, of Table 2.4, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 1, 2, 4, 8, 9, 10, 11, 12, 14,15, 16, 20, 23, 24, 25, 28, 29, 31,33, 35,37, 39, 41, 42, 43, 44, 45, 46, 49, and/or 50 of Table 2.4

is indicative for the presence of ball when ball is distinguished from pret,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21,

22, 23, 24, 25, 26, 27, 28, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table of Table 2.5, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 29, 30 and/or 39, of Table 2.5,

is indicative for the presence of ball when ball is distinguished from prob,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 7, 9, 10, 11, 13, 17, 18, 21, 24, 25, 27, 29, 30, 31, 36, 37, 38, 40, 42, 43, 45, 46, 49, and/or 50 of Table 2.6, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 6, 8, 12, 14, 15, 16, 19, 20, 22, 23, 26, 28, 32, 33, 34, 35, 39, 41, 44, 47, and/or 48 of Table 2.6,

is indicative for the presence of cpre when cpre is distinguished from cpreph,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 4, 5, 6, 8, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 24, 25, 27, 28, 29, 30, 31, 32, 35, 36, 38, 40, 41, 43, 44, 45, 46, 48, 49, and/or 50 of Table 2.7, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 3, 7, 9, 11, 22, 26, 33, 34, 37, 39, 42, 47, of Table 2.7,

is indicative for cpre when cpre is distinguished from kort,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 20, 28, 31, 37, 38, and/or 50 of Table 2.8, and/or

a higher expression of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 29, 30, 32, 33, 34, 35, 36, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, and/or 49 of Table 2.8

22, 23, 24, 25, 27, 28, 29, 30, 31, 32, 34, 35, 36, 37, 38, 39, 40, 42, 43, 44, 45, 46, 47, 48, and/or 50 of Table 2.9,

a higher expression of at least one polynucleotide defined by any of the numbers 26, 33, 41, and/or 49 of Table 2.9

is indicative for cpre when cpre is distinguished from prob,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 3, 6, 12, 17, 23, 28, 34, 35, and/or 41, of Table 2.10, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 1, 2, 4, 5, 7, 8, 9, 10, 11, 13, 14, 15, 16, 18, 19, 20, 21, 22, 24, 25, 26, 27, 29, 30, 31, 32, 33, 36, 37, 38, 39, 40, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table 2.10

is indicative for cpreph when cpreph is distinguished from kort,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 42, and/or 43, of Table 2.11, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 44, 45, 46, 47, 48, 49, and/or 50 of Table 2.11,

is indicative for cpreph when cpreph is distinguished from pret,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 3, 5, 8, 9, 11, 12, 13, 15, 18, 21, 24, 27, 28, 29, 32, 34, 36, 38, 41, 42, 43, 46, 47, 48, of Table 2.12, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 2, 4, 6, 7, 10, 14, 16, 17, 19, 20, 22, 23, 25, 26, 30, 31, 33, 35, 37, 39, 40, 44, 45, 49, and/or 50 of Table 2.12

is indicative for cpreph when cpreph is distinguished from prob

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 19, and/or 40, of Table 2.13

a higher expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table 2.13, .

is indicative for kort when kort is distinguished from pret,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 4, 7, 9, 10, 11, 13, 14, 15, 16, 17, 20, 21, 22, 28, 29, 31, 32, 33, 35, 36, 37, 40, 41, 42, 43, 45, 47, 48, and/or 50 of Table 2.14, and/or

a higher expression of at least one polynucleotide defined by any of the numbers 2, 3, 5, 6, 8, 12, 18, 19, 23, 24, 25, 26, 27, 30, 34, 38, 39, 44, 46, and/or 49, of Tabl 2.14

is indicative for kort when kort is distinguished from prob,

and/or wherein

a lower expression of at least one polynucleotide defined by any of the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, and/or 50 of Table 2.15,

is indicative for pret when pret is distinguished from prob.

2. The method according to claim 1 wherein the polynucleotide is labelled.

3. The method according to claim 1 or 2, wherein the label is a luminescent, preferably a fluorescent label, an enzymatic or a radioactive label.

4. The method according at least one of the claims 1-3, wherein the expression level of at least one, preferably of at least [illegible], more preferably of at least [illegible], most [illegible]

5. The method according to at least one of the claims 1-4, wherein the expression level of markers expressed lower in a first subtype than in at least one second subtype, which differs from the first subtype, is at least 5 %, 10% or 20%, more preferred at least 50% or may even be 75% or 100%, i.e. 2-fold lower, preferably at least 10-fold, more preferably at least 50-fold, and most preferably at least 100-fold lower in the first subtype.

6. The method according to at least one of the claims 1-4, wherein the expression level of markers expressed higher in a first subtype than in at least one second subtype, which differs from the first subtype, is at least 5 %, 10% or 20%, more preferred at least 50% or may even be 75% or 100%, i.e. 2-fold higher, preferably at least 10-fold, more preferably at least 50-fold, and most preferably at least 100-fold higher in the first subtype.

7. The method according to at least one of the claims 1-6, wherein the sample is from an individual having ALL.

8. The method according to at least one of the claims 1-7, wherein at least one polynucleotide is in the form of a transcribed polynucleotide, or a portion thereof.

9. The method according to claim 8, wherein the transcribed polynucleotide is a mRNA or a cDNA.

10. The method according to claim 8 or 9, wherein the determining of the expression level comprises hybridizing the transcribed polynucleotide to a complementary polynucleotide, or a portion thereof, under stringent hybridization conditions.

11. The method according to at least one of the claims 1-7, wherein at least one polynucleotide is in the form of a polypeptide, or a portion thereof.

12. The method according to at least one of the claims 8, 9 or 12, wherein the determining of the expression level comprises contacting the polynucleotide or the polypeptide with a compound specifically binding to the polynucleotide or the polypeptide.

13. The method according to claim 12, wherein the compound is an antibody, or a fragment thereof.

14. The method according to at least one of the claims 1-13, wherein the method is carried out on an array.

15. The method according to at least one of the claims 1-14, wherein the method is carried out in a robotics system.

16. The method according to at least one of the claims 1-15, wherein the method is carried out using microfluidics.

17. Use of at least one marker as defined in at least one of the claims 1-3 for the manufacturing of a diagnostic for distinguishing immunologically defined ALL subtypes Pro-B-ALL, c-ALL, Pre-B-ALL, c-ALL/Pre-B-ALL, mature B-ALL, precursor B-ALL, Pro-T-ALL, Pre-T-ALL, cortical T-ALL, mature T-ALL, and/or T-ALL.

18. The use according to claim 17 for distinguishing immunologically defined ALL subtypes Pro-B-ALL, c-ALL, Pre-B-ALL, c-ALL/Pre-B-ALL, mature B-ALL, precursor B-ALL, Pro-T-ALL, Pre-T-ALL, cortical T-ALL, mature T-ALL, and/or [illegible]

c-ALL, Pre-B-ALL, c-ALL/Pre-B-ALL, mature B-ALL, precursor B-ALL, Pro-T-ALL, Pre-T-ALL, cortical T-ALL, mature T-ALL, and/or T-ALL, in combination with suitable auxiliaries.

20. The diagnostic kit according to claim 19, wherein the kit contains a reference for the immunologically defined ALL subtypes Pro-B-ALL, c-ALL, Pre-B-ALL, c-ALL/Pre-B-ALL, mature B-ALL, precursor B-ALL, Pro-T-ALL, Pre-T-ALL, cortical T-ALL, mature T-ALL, and/or T-ALL.

21. The diagnostic kit according to claim 20, wherein the reference is a sample or a data bank.

22. An apparatus for distinguishing immunologically defined ALL subtypes Pro-B-ALL, c-ALL, Pre-B-ALL, c-ALL/Pre-B-ALL, mature B-ALL, precursor B-ALL, Pro-T-ALL, Pre-T-ALL, cortical T-ALL, mature T-ALL, and/or T-ALL in a sample containing a reference data bank.

23. The apparatus according to claim 22, wherein the reference data bank is obtainable by comprising
    (a) compiling a gene expression profile of a patient sample by determining the expression level of at least one marker selected from the markers identifiable by their Affymetrix Identification Numbers (affy id) as defined in Tables 1, and/or 2, and
    (b) classifying the gene expression profile by means of a machine learning algorithm.

24. The apparatus according to claim 23, wherein the machine learning algorithm is selected from the group consisting of Weighted Voting, K-Nearest Neighbors, Decision Tree Induction, Support Vector Machines, and Feed-Forward Neural Networks, preferably Support Vector Machines.

25. The apparatus according to at least one of the claims 22-24, wherein the apparatus contains a control panel and/or a monitor.

26. A reference data bank for distinguishing immunologically defined ALL subtypes Pro-B-ALL, c-ALL, Pre-B-ALL, c-ALL/Pre-B-ALL, mature B-ALL, precursor B-ALL, Pro-T-ALL, Pre-T-ALL, cortical T-ALL, mature T-ALL, and/or T-ALL obtainable by comprising
    (a) compiling a gene expression profile of a patient sample by determining the expression level of at least one marker selected from the markers identifiable by their Affymetrix Identification Numbers (affy id) as defined in Tables 1, and/or 2, and
    (b) classifying the gene expression profile by means of a machine learning algorithm.

27. The reference data bank according to claim 26, wherein the reference data bank is backed up and/or contained in a computational memory chip.

F. Hoffmann-La Roche AG
Roche Diagnostics GmbH

November 04, 2003
R62510EP BÖ/AMS

Abstract

Disclosed is a method for distinguishing immunologically defined ALL subtypes Pro-B-ALL, c-ALL, Pre-B-ALL, c-ALL/Pre-B-ALL, mature B-ALL, precursor B-ALL, Pro-T-ALL, Pre-T-ALL, cortical T-ALL, mature T-ALL, and/or T-ALL in a sample by determining the expression level of markers, as well as a diagnostic kit and an apparatus containing the markers.